

What is claimed is:

1. A system comprising:

5 an event detection server in communication with a social media platform over a network, the event detection server including a processor and memory storing instructions that, in response to receiving in real-time social media data from the social media platform, cause the processor to:

10 apply a set of filter modules to the social media data to generate a data store, the data store comprising a set of identified concepts and corresponding attributes of the social media data;

15 select one of the set of identified concepts from the data store using a corresponding threshold value associated with the attributes of the social media data, wherein the corresponding threshold value is associated with three or more distinct attributes of the social media data related to the identified concept and wherein one of the attributes of the social media data is an authorship value and the corresponding threshold value represents at least three or more similar identified concepts associated with different authorship values;

generate an event detected cluster using the selected identified concept;

20 generate a verification score for each item of social media data associated with the event detected cluster, the verification score being indicative of a veracity of the event detected cluster, wherein the verification score is determined by analyzing at least a user category, a social media category and an event features category relating to each item of social media data associated with the event detected cluster; and

25 present the event detected cluster and verification score to the user on a graphical user interface,

wherein each of the set of identified concepts has a predefined time frame to meet the corresponding threshold value and be selected such that if the corresponding threshold value has not been met within a time period that does not exceed the predefined time frame, the processor is caused to re-evaluate the respective identified concept with newly received social media data and their corresponding attributes, and if the corresponding threshold value has not been met within a time period that exceeds the predefined time frame, the processor is caused to remove the identified concept from the data store, and

wherein once the event detected cluster is generated, the processor is caused to remove from the data store the selected identified concept and its corresponding attributes of the social media data.

2. The system of claim 1, wherein one of the set of filter modules detects language of the social media data and deleting the social media data that is not in English.

3. The system of claim 1 or 2, wherein one of the set of filter modules detects profanity used in the social media data and removes the social media data containing the detected profanity.

4. The system of any one of the preceding claims, wherein one of the set of filter modules detects at least one of spam, chat and advertisement in the social media data and removing the social media data that contains the at least one detected spam, chat and advertisement.

5. The system of any one of the preceding claims, wherein one of the set of filter modules applies Parts-Of- Speech tagging of the social media data.

5 6. The system of any one of the preceding claims, wherein one of the set of filter modules analyzes semantic and syntactic structures in the social media data to determine identified concepts in the social media data.

10 7. The system of any one of the preceding claims further comprising a topic categorization module configured to generate a topic classification for the event detected cluster.

15 8. The system of any one of the preceding claims further comprising a summary module configured to generate a summary for the event detected cluster.

20 9. The system of any one of the preceding claims further comprising a newsworthiness module configured to generate a newsworthy score for the event detected cluster.

25 10. The system of any one of the preceding claims further comprising an opinion module configured to identify opinion or fact for each item of social media data associated with the event detected cluster.

30 11. The system of any one of the preceding claims further comprising a credibility

module configured to generate a credibility score for each item of social media data associated with the event detected cluster.

5 12. The system of any one of the preceding claims, wherein the user category comprises at least one of name of author, description of author, URL of author, location of author, location of the author matching the location of the event, author being a witness to the event, protection level of the author's account, or verification of the author.

10

13. The system of any one of the preceding claims, wherein the social media category comprises at least one of multimedia, url, elongated word, url from news source, or word sentiment associated with the social media data.

15

14. The system of any one of the preceding claims, wherein the event features category comprises at least one topic of the event and portion of the social media that deny, believe or question the event associated with each item of the social media data

20

15. The system of any one of the preceding claims, wherein the social media category is twitter data and the event features category further comprises at least one of a count of the most retweeted tweets, a frequency of retweeted tweets or a
25 frequency of hashtags associated with each item of the social media data.

16. A computer-implemented method for detecting and verifying an event from social media data, the method comprising:

receiving in real-time, by an event detection server in communication with a social media platform over a network, social media data from the social media platform, the event detection server including a processor and memory instructions;

applying, by the event detection server, a set of filter modules to the social media data to generate a data store, the data store comprising a set of identified concepts and corresponding attributes of the social media data;

selecting, by the event detection server, one of the set of identified concepts from the data store using a corresponding threshold value associated with the attributes of the social media data, wherein the corresponding threshold value is associated with three or more distinct attributes of the social media data related to the identified concept and wherein one of the attributes of the social media data is an authorship value and the corresponding threshold value represents at least three or more similar identified concepts associated with different authorship values;

generating, by the event detection server, an event detected cluster using the selected identified concept;

generating, by the event detection server, a verification score for each item of social media data associated with the event detected cluster, the verification score being indicative of a veracity of the event detected cluster, wherein the verification score is determined by analyzing at least a user category, a social media category and an event features category relating to each item of social media data associated with the event detected cluster; and

presenting the event detected cluster and verification score to the user on a graphical interface,

wherein each of the set of identified concepts has a predefined time frame to meet the corresponding threshold value and be selected such that if the corresponding threshold

value has not been met within a time period that does not exceed the predefined time frame, the processor of the event detection cluster is caused to re-evaluate the respective identified concept with newly received social media data and their corresponding attributes, and if the corresponding threshold value has not been met within a time period that exceeds the predefined time frame, the processor is caused to remove the identified concept from the data store, and

wherein once the event detected cluster is generated, the processor is caused to remove from the data store the selected identified concept and its corresponding attributes of the social media data.

17. The method of claim 16, wherein the user category comprises at least one of name of author, description of author, url of author, location of author, location of the author matching the location of the event, author being a witness to the event, protection level of the author's account, or verification of the author.

18. The method of claim 16 or 17, wherein the social media category comprises at least one of multimedia, url, elongated word, url from news source, or word sentiment associated with the social media data.

19. The method of any one of claims 16 to 18, wherein the event features category comprises at least one topic of the event and portion of the social media that deny, believe or question the event associated with each item of the social media data.

20. The method of any one of claims 16 to 19, wherein the social media category is

twitter data and the event features category further comprises at least one of a count of the most retweeted tweets, a frequency of retweeted tweets or a frequency of hashtags associated with each item of the social media data.